MEMORANDUM

To: National Emission Inventory (NEI) Data Preparers
From: Anne Pope, U.S. EPA, Emission Factor and Inventory Group
Re: NEI Quality Assurance and Data Augmentation Steps for Point Sources

Background

For use in air quality modeling exercises by EPA and others, the National Emission Inventory (NEI) point emission files need to be converted to inputs suitable for running through various emission processors (SMOKE, EMS-2001, EMS-HAP, etc.). More information on emissions modeling is located at www.epa.gov/ttn/chief/emch/. In addition to converting the formats into files that can be read by the emissions processors, the Emission Factor and Inventory Group (EFIG) runs a list of quality assurance routines to try and catch any missing or "out-of-range" parameters which may cause issues in the processor runs. Some of these routines will result in modifications to the data. The purpose of this memo is to describe and document the augmentation for data with missing or out-of-range values. Over time, as we prepare the NEI for use in air quality exercises, we will continue to improve and document these data processing routines.

This memo describes how we will modify the emission inventory files you provide us, in the interest of respectful data use. The NEI files provided on the EPA FTP site clearly identify all records that have defaulted parameters and document how missing or out-of-range parameters have been defaulted. I hope that by understanding the consequences of submitting files with missing or "out-of-range" parameters, your agency can better plan and prioritize its emissions inventory development activities.

QA Routines

Emission Release Point - Physical Parameters

Because of the detail in emissions processing, valid parameters for the physical characteristics of each release point (stack height, diameter, temperature, velocity, and flow) are necessary to correctly place facility release points and associated emissions into vertical layers for proper air quality modeling. However, it is frequently noted that not all of the physical characteristics of each release point are reported in the data submittals. Sometimes there are values reported but they are physically implausible, suggesting that a misunderstanding regarding the meaning of the data field, the units of measure, or a transcription of data has occurred.

In the National Emission Input Format (NIF), the following emission release point types are reported. (The address for information on the NIF is: www.epa.gov/ttn/chief/nif/index/.)

Fugitive	Vertical
Horizontal	Goose Neck
Vertical with Rain Cap	Downward Facing Vent
Unknown	

We assume that only emission release point types reported as fugitive are fugitive. We assume that all other emission release point types are non-fugitive. For the 2002 NEI, we will evaluate the SCCs of emission release point types reported as fugitive to determine if the emission release point type is reported correctly as a fugitive.

We employ a routine that compares each emission release point parameter to a minimum and maximum range of values and when that parameter is missing or is found to exist outside of that range we modify the parameter. We also check non-fugitive stack parameters for internal consistency. The following steps describe the process of finding and replacing missing, out-of-range, or internally inconsistent stack parameters.

Step 1: For fugitive emission release points, replace stack parameters
For fugitive emission release points, we first compare the existing height against the following range thought to be representative of the minimum and maximum values allowable for most fugitive emission release points.

Fugitive Release Height (ft):	0.1 to 100

If the height is valid, we keep the height and replace all other stack parameters with the defaulted values listed below. If the height is invalid, we replace all stack parameters with the defaulted values. NOTE: Ideally for a fugitive emission release point, the defaulted stack height would be one half the building height. In addition an emissions processor would be able to distribute the fugitive emissions appropriately among the layers for grid modeling. Currently this is not available. Thus at this time, we recommend using the following defaulted values.

Stack Height (ft):	10
Stack Temperature (^{o}F):	72
Stack Diameter (ft):	0.003
Stack Velocity (ft/sec):	0.0003
Stack Flow (cu ft/sec):	0

Step 2: For non-fugitive emission release points, find out-of-range or missing stack parameters or inconsistency in stack parameters
For non-fugitive emission release points, we first compare existing stack parameters against a set of the following ranges thought to be representative of the minimum and maximum values allowable for most emission release points.

Stack Height (ft):	0.1 to 1,000
Stack Temperature (^{o}F):	50 to 1,800
Stack Diameter (ft):	0.1 to 50
Stack Velocity (ft/sec):	0.1 to 560
Stack Flow (cu ft/sec)	0.001 to 1,100,000

Then we determine any inconsistencies in stack parameters by conducting the following two steps.

A. For stack diameter, we compare the stack diameter to the stack height. For non-fugitive emission release points, the stack height may not be less than stack diameter.

B. We determine the internal consistency between diameter, velocity and flow rate using the following equation.

Stack Flow [cu ft/sec] = (Π [Pi] * (Stack Diameter [ft] / 2) ^ 2) * Stack Velocity [ft/sec]

If the calculated flow and the reported flow are within 10 % of one another, then internal consistency is assumed to be valid.

If all parameters are found to exist within the bounds of the emission release point ranges, and the consistency checks (A) and (B) are satisfied, no additional steps are taken.

If any parameter is missing or out-of range, or if the parameters fail the internal consistency tests, the parameter is replaced using the established procedures described in Step 3.

Step 3: Replace stack parameters for non-fugitive emission release points

The first step in replacing stack parameters is to determine if there are problems with stack height or diameter. Because stack height and diameter are the physical parameters that are most easily measured or estimated, when there are problems with these parameters, then the entire set of stack parameters are deemed questionable. If either height or diameter is missing or out-of range, or if the stack diameter is greater than stack height, then all 5 parameters are defaulted using national default sets of physical parameter data. No additional steps are taken once all 5 parameters are defaulted.

If stack height and diameter do not need replacement, then velocity and flow are evaluated next. Because of the relational nature of velocity, flow and diameter, Attachment A is provided to explicitly describe all of the possible cases of data replacement within those relationships.

Finally, in cases where all 5 parameters have not been defaulted, and velocity and flow have been evaluated and replaced if necessary, temperature is evaluated. If temperature is missing or out-of-range, then the temperature is defaulted using national default sets of physical parameter data.

National default sets of physical parameter data organized by SCC and SIC are used to replace all 5 parameters by the following hierarchy.

1. SCC match
2. facility level SIC match
3. national default for release points, if no SCC or SIC match is possible

From the previously released version of the 1999 NEI (currently NEI 1999 v.1), default look-up tables are generated by SCC and SIC, to report the average value calculated for stack height, diameter, temperature and velocity for all emission release point types that are coded as stacks. Only non-zero values are used in the averaging. The stack flow is calculated by using the average diameter and average velocity and the equation in Step 2 B. Separate look-up tables are

prepared for SCC and SIC. These files can be found at the following address.
www.epa.gov/ttn/chief/emch/invent/

In the actual data set where an out-of-range parameter value has been found for a specific stack, all of the SCCs or SICs characterizing emissions through that stack are first determined. In the case of multiple SCCs or SICs, a match is done for each in the respective look-up table to find a default replacement value for the out-of-range parameter. When there are multiple default replacement values by SCC / SIC possible for a specific stack, we used the default record having the lowest stack height to modify and replace that out-of-range release parameter.

If no SCC or SIC match is possible, we used the following national default values for the stack parameters.

Stack Height (ft):	10
Stack Temperature (^{o}F):	72
Stack Diameter (ft):	1
Stack Velocity (ft/sec):	15
Stack Flow (cu ft/sec):	12

Emission Release Point - Geographical Location

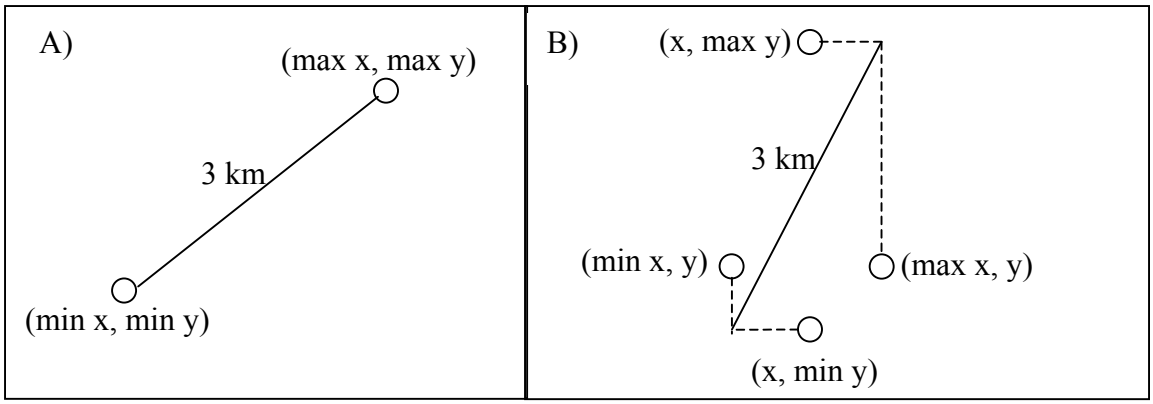
Latitude and longitude are needed to correctly place facility emission release points and associated emissions into specific (grid cell, census tract, etc.) geographic domains for proper emissions modeling. Many instances have been reported where FIPS state and county codes do not correspond to the latitude and longitude values and/or to the zip code supplied with each facility.

We use a routine to assess the validity of the latitude and longitude values, to replace values if necessary, and to fill-in missing data points. The stages of the routine are described below. All latitude/longitude data points are assumed to be reported in units of decimal degrees.

Step 1. Find and replace the latitude/longitude of emission release points within a facility that are located at distances greater than 3.0 km of other release points in the facility.
This step includes determining if an individual release point within a facility is within 3.0 km of other emission release points. (i.e., to ensure no stacks within a single facility are located miles apart).

If there is only one emission release point, the process is complete. Proceed to Step 2.

If there are more than one emission release points, we identify the emission release points with the highest and the lowest latitudes and emission release points with the highest and lowest longitudes at the site. We calculate the distances between the highest and lowest latitudes and between the highest and lowest longitudes. Please refer to the diagram below for more details on how we conduct this analysis.



This method may but does not necessarily measure the true distance between emission release points. For example, if a facility has 2 emission release points, the true distance is measured by examining minimum and maximum latitudes/longitudes at all emission release points described by Example A in the diagram. If, for example, a facility has 4 emission release points as in Example B of the diagram, the distance calculated does not represent the distance between any two specific emission release points, but rather defines the maximum distance that *could* exist between any two emission release points given all of the points in the set (Example B). This analysis identifies sites whose emission release points are potentially far apart and whose coordinates need correction, but it is not designed to identify the specific outlying emission release points.

A. If the greatest distance between latitudes and between longitudes is less than 3.0 km, the process is complete. Proceed to Step 2.

B. If the distances between the highest and lowest latitudes or the highest and lowest longitudes are greater than 3.0 km, the latitudes and longitudes of all emission release points within a facility are evaluated. If there are records whose latitudes and longitudes are found to be at a distance of more than 3.0 km, the SIC or NAICS codes are examined to determine if the distance between emission release points is technically correct.

- For a source category, if it is acceptable that the distance between emission release points is greater than 3.0 km, the process is complete. Proceed to Step 2.

- For a source category, if it is determined that the distance between emission release points should not be more than 3.0 km, then the following steps are conducted.

1. The distances between all emission release points within the facility is calculated and outlier(s) are identified. For example if four emission release points are present (A, B, C and D), then the distances between the latitudes and longitudes of A & B, A & C, A & D, B & C, B & D, and C & D are calculated.

2. An average site latitude/longitude is calculated using only the acceptable coordinates. This average site latitude/longitude is then used to replace the inaccurate latitude/longitude values. After this step, the process is complete and all the latitude/longitude data are

assumed to be correct. Proceed to Step 2.

Step 2. Find and replace latitude/longitude of emission release points that are out-of-county boundary or missing

This step includes the use of a GIS overlay to plot each latitude/longitude value and compare it to the physical boundaries of the FIPS county to which the value is associated (i.e., to ensure no stacks are located in the oceans or in far away states). Detailed county boundaries using a scale of 1 to 100,000 or better are used in the GIS overlay.

If the plotted release point is within ten kilometers of the county, the point is assumed to be valid and neither latitude/longitude nor county FIPS are corrected. The process is complete.

If the plotted release point is found to exist more than ten kilometers outside of the county or if the latitude/longitude is missing, then the latitude/longitude of each emission release point is replaced using the following hierarchy. Only one method is used for missing or out-of-boundary latitude/longitude within a facility, i.e., method A is not used for one emission release point at a facility and method D for another emission release point at the same facility.

A. Use Facility Specific Data - A check is first completed to see if there are any other emission release points at the facility that exist within the ten kilometer zone. The county FIPS code for the emission release points that are outside the ten kilometer boundary is compared with county FIPs codes of other release points within the facility that exist within the ten kilometer boundary. If the other valid emission release points within a facility have a different county FIPS from the emission release points that are in question, then the county FIPS is changed and the latitude/longitude of the emission release points rechecked to see if the latitude/longitude of emissions release points in question are now valid.

If any emission release points in the facility exist within the ten kilometer boundary, an average site location is estimated using the latitude/longitude of the group of emission release points located within the ten kilometer zone. The county FIPS of the group of emission release points within the ten kilometer zone and the average site latitude/longitude are used to replace those latitude/longitude values and county FIPS found to exist at emission release points outside of the ten kilometer boundary. The process is complete.

B. *Use Geocoding software* – If none of the reported emission release point latitude/longitude values exist within the ten kilometer boundary of the county according to the facility's FIPS code, geocoding software is used. More information on Tele Atlas North America "EZ-Locate" geocoding software is located at www.geocode.com/.

The first step in using Geocoding software is to check the quality of the zip code provided in the inventory with a zip code QA file. The zip code file can be found at the following address, www.epa.gov/ttn/emch/invent/. The source of the zip code file is ESRI and the file uses zip code information from the US Postal Service and FIPS code information from the US Department of Commerce, National Institute of Information and Technology. Zip codes

and FIPS codes reported in the Site records are compared with zip codes and FIPS codes in the zip code file. The zip code file is then used to replace zip codes and FIPS codes for all records needing correction prior to using geocoding software.

After correcting zip codes, a list of sites that have emission releases points with missing or erroneous latitude/ longitude is compiled. The file contains site name, the physical addresses of the sites, state FIPS and county FIPS. The file is submitted to the geocoding software. Using the inventory record's street address, the geocoding software matches using the following hierarchy.

1. First the software standardizes the street address and looks for an exact address match.
2. If an exact match cannot be found, the software then tries to match at the single street block, as defined by Geocoder's documentation. The latitude/longitude are located at the centroid of the single street block.
3. If a single street block match cannot be found, then the software tries to match on the 5-digit zip code plus 2 digits. The latitude/longitude are located at the centroid of the 5 digit zip code plus 2 digits.
4. If a match to 5-digit zip code plus 2 digits cannot be found, then the software tries to match to the 5-digit zip code alone. The latitude/longitude are located at the centroid of the 5-digit zip code.
5. If a match to 5-digit zip code cannot be found, then the software tries to match to 3-digit zip code. The latitude/longitude are located at the centroid of the 3- digit zip code.
6. If a match to a 3-digit zip code cannot be found, then the software provides an "ambiguous" match which is a match to multiple non-standardized street segments.

If the geocoding software finds valid latitude/longitude data, the process is considered to be complete and all latitude/longitude are assumed to be valid. The county FIPS code will be changed if necessary to match the geocoded county FIPS. The process is complete.

C. *Use FRS database of EPA plant information* - The Facility Registry System (FRS) is a centrally managed database that identifies facilities, sites or places subject to environmental regulations or of environmental interest. The latitude/longitude data in FRS represents the general site location. The facilities can be matched by FRS ID, name, FIPS and address. More information on the EPA Facility Registry System (FRS) is located at: www.epa.gov/enviro/html/facility.html.

If the identified county in FRS for a facility does not match the reported county in the inventory, the FRS data should not be used. If the FRS finds valid latitude/longitude data with corresponding valid county, the process is considered complete and all the latitude/ longitude data are assumed to be valid.

D. *Site Release Point at County Centroid* - If, after each of these stages, an emissions release point latitude/longitude data set is still found to be missing or invalid, site the emission release

point to the county centroid.

Operating Parameters

In addition to the correct placement of emissions, air quality modeling attempts to represent the actual physical and chemical processes as they occur over a specific duration of time. As such, it is important that the temporal allocation of emissions be as accurate as possible. Temporal allocation can be thought of as an accounting of emissions variation over time. The simplest temporal allocation is for a steady-state emissions source that continually releases emissions at the same rate all the time. Under actual conditions, however, steady-state emission sources are quite rare. Instead, emissions sources may operate only in the winter, not operate on Sundays, or their activity may peak during certain hours of the day.

Temporal allocations based on seasonal fuel throughput percentages, weeks per year, days per week, and hours per day allow emissions variability to be correctly modeled during the desired modeling periods. An example would be an emission source reporting activity for fifty-two weeks per year, five days a week, and eight hours a day. Under the current temporal profile assignment paradigm, this source would be allocated a temporal emission release between the hours of 9 AM and 5 PM, Monday through Friday, each week of the year.

At this time, inventory reported operating parameters are not used to assign emissions to month, week, day, or hour. Instead, SCC-based temporal profiles are assigned to represent when emissions from source types typically occur. Documentation of the procedures and allocation profiles currently used in this allocation can be found on EFIG's Emissions Modeling Clearinghouse website. (www.epa.gov/ttn/chief/emch/temporal)

It is the intent of EPA to eventually use actual reported operating parameters to assign when individual sources release their emissions. Currently, however, many instances exist where provided data are outside of acceptable ranges or missing completely and result in the assignment of a flat profile of twenty four hour a day, seven days a week, fifty two weeks a year operation (i.e., the same emission release each hour of each day of each year). Having these operating parameters provided with the emissions inventory will eventually result in a better, more accurate, source specific distribution of emissions on a diurnal, daily, weekly, and seasonal level.

Annual and Episodic Emissions Check

A routine will be used to check each segment's pollutants for cases of episodic emissions exceeding annual emission values. Because segment-level emissions may be reported and/ or calculated for different time periods, e.g., annual, daily, seasonal, etc., and for the same or different pollutants, verification is needed to ensure that episodic emissions are not higher than annual emissions.

When episodic emissions are found to exceed annual emission estimates, the episodic value is recalculated using annual emissions and the operating throughput parameters validated in the previous step.

If episodic emissions are reported as positive and annual emissions are reported as zero, we contact the state or local agency or tribe to determine if the episodic value for the reporting period specified is also the annual emission estimate (ex., snow mobile emissions reported for months of January – March). If the agency or tribe indicates that the episodic emission estimate for a source category is not the annual emission estimate, then the episodic emission estimate and operating throughput parameters are used to estimate annual emissions.

We will be revising the augmentation procedure for generating annual emission estimates when episodic emission estimates are reported and annual emissions are not reported in the 2002 NEI. We will also be adding a data field to distinguish between annual and episodic emission estimates in the 2002 NEI as well as flags to indicate the methodology of defaulting annual and episodic emission estimates.

Attachment A - Stack Parameter Data Replacement Matrix

X = Data value is present

Diameter	Velocity	Flow	Action
X	X	X	1. Check that velocity and flow are within range. A. If velocity is not within range and flow is within range: Calculate velocity using formula in QA memo Check that calculated velocity is within range. If not: Default all 5 parameters using approach defined in Step 3 of QA memo. B. If velocity is within range and flow is not within range: Calculate flow using formula in QA memo. Check that calculated flow is within range. If not: Default all 5 parameters using approach defined in Step 3 of QA memo. C. If velocity and flow are not in range, then: Default all 5 parameters using approach defined in Step 3 of QA memo.
X	-	X	1. Check that flow is within range. If not, then: Default all 5 parameters using approach defined in Step 3 of QA memo. No additional steps are taken if all 5 parameters are defaulted. 2. Calculate velocity using formula in QA memo. 3. Check that calculated velocity is within range. If not: Default all 5 parameters using approach defined in Step 3 of QA memo.
X	X	-	1. Check velocity is within range. If not, then: Default all 5 parameters using approach defined in Step 3 of QA memo. No additional steps are taken if all 5 parameters are defaulted. 2. Calculate flow using formula in QA memo. 3. Check that calculated flow is within range. If not: Default all 5 parameters using approach defined in Step 3 of QA memo.
X	-	-	1. Default velocity using approach defined in Step 3 of QA memo. 2. Calculate flow using formula in QA memo. 3. Check that calculated flow is within range. If not: Default all 5 parameters using approach defined in Step 3 of QA memo.